Business models for Linked Open Government Data: what lies beneath?

Authors of the study: Phil Archer – W3C, Makx Dekkers – AMI Consult, Stijn Goedertier – PwC EU Services, Nicolas Hazard – PwC EU Services, Nikolaos Loutas – PwC EU Services
Project Officers: Athanasios Karalopoulos, Vassilios Peristeras, Susanne Wigard – EU Commission

This paper lays out the base for a plenary presentation about a study over Business models for Linked Open Government Data.

Abstract

This presentation discusses a study on business models for Linked Open Government Data (LOGD). This study was performed between April and August 2013 and commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission. Some public organisations decided to investigate the use of Linked Open Governance Data as a means of providing machine readable-format. This study investigated this topic from a business model point of view.

The study provides a theoretical framework that can be used to study business models for linked open government data. In addition, the study provides interesting conclusions based on the analysis of 14 case studies. The prevalent business model that emerged is the one where the investment and maintenance costs of a LOGD service are covered through ongoing public funding. In all cases, the LOGD service is provisioned free of charge. Many providers do not yet have a clear view of the consumers of their data. In the study, we have not seen a wide reuse by third parties. This may be because providers do not yet provide operational guarantees or because Linked Data requires acquisition of new skills on the side of the reusers. However, as many data providers are in the process of producing massive amounts of LOGD and given the opportunities for efficient data integration that Linked Data offers, more reusers are expected to find their way to the data.

Keywords: Open Data, PSI, European Union, Linked Open Government Data

Introduction

Open Government is supported by the availability of Open Government Data. Open Government Data refers to data available under an open licence produced or commissioned by governments or government controlled entities which can be used, re-used and redistributed by anyone and for free. Open data is key to achieve a number of goals both from a data publisher and a data re-user perspective, which could include transparency, research or the creation of business opportunities.

This presentation discusses the outcome of a study on Linked Open Government Data (LOGD) performed by the ISA Programme between April and August 2013. The study addressed the following research topics:

1. the value of LOGD for businesses, citizens, and public administrations,
2. the cost structures behind the provision of LOGD,
3. the revenue streams linked to the consumption of LOGD services, and

4. enablers and barriers with regard to the value creation of LOGD.

The presentation consists of three main parts. First we discuss the theoretical framework that we used. Second, we provide an overview of the 14 case studies on LOGD that we analysed in the study. Finally, we summarise the findings, the identified enablers and barriers.

Business models for Linked Government Data: what lies beneath?

The theoretical framework used is structured according to the nine areas in the Business Model Canvas [Osterwalder]. In the LOGD ecosystem, public administrations are data providers that provide Open Government Data as an online LOGD service to data consumers – citizens, businesses and other public administrations. Instead of downloading and processing a whole dataset, LOGD allows a data consumer to retrieve specific information about the entity of his interest, by resolving its Web identifier (URI). The data is provided in different machine-readable formats, ready to be linked and meshed-up with other data.

The theoretical framework hence puts the following value proposition of LOGD forward:
1. LOGD offers flexible data integration;
2. LOGD leads to an increase in data quality;
3. the use of LOGD gives rise to new services; and
4. LOGD reduces data integration costs.

To enable this value proposition, LOGD providers must have a URI policy that lays down the expected service levels of the Linked Data service; long-term persistence being one of the most important service levels. Governments should make their URI policy explicit, so that LOGD consumers can rely on LOGD services with confidence and other data providers can link to these URIs or reuse these URIs to denote identical concepts. The use of URIs as common identifiers to identify identical concepts in disparate datasets is a prerequisite to unlock the positive network effects of LOGD. The theoretical framework also provides a number of favourable conditions under which public administrations could consider providing LOGD:
- **Nature of the data**: there are no restrictions (e.g. no personal data protection and/or privacy concerns).
- **Positive network effects**: the publication of LOGD can reduce the costs of resolving interoperability conflicts in information exchange and ease data integration. This may be particularly the case for data models and reference data that is used by many in different contexts.
- **No other economic agent wants to/can offer it:** in case no other economic agent wants to or can offer some reference data as LOGD, it may make sense for governments to assume this task, e.g. in the case of Base Registers. Otherwise, providing the LOGD service would compete with the services of other market players.
- **Economies of scale:** governments may already have the infrastructure in place to provide the service and can provide LOGD with little additional costs.
- **Guarantees of stability and persistence:** governments are in a good position to guarantee stability and persistence of the LOGD service.

Thirty seven cases are identified in which public administrations have used LOGD to make open government data available as a service on the Web; fourteen case have been selected for further analysis according to the aforementioned theoretical framework. The information that was gathered for each case study, both by conducting an interview with the stakeholders and by performing desk research.

We summarise below the main findings of the case studies, according to the nine areas of the Business Model Canvas:

- **Value proposition**: the main driver for the use of LOGD in the cases investigated is that it allows for flexible data integration; this helps to increase data quality by allowing cross-references to authoritative data to be included and may drive future development of new services. The use of LOGD increases the efficiency of the internal operation of the provider and allows them to fulfil their public task more effectively and efficiently Whether this has given rise to new services or actually reduces costs varies from one case to another. The case studies do not provide evidence that LOGD has contributed to increased data quality, for example via self-service or crowd-sourcing mechanisms. The case studies also reveal that little effort is currently spent on quantitatively measuring the usage and benefits of LOGD. This may be put down to the fact that case study participants have implemented LOGD firstly for their internal data consumption, and are not yet in contact with external consumers.
- **Key resources**: LOGD is applied most successfully in reference data; URI design policies are generally in place, while persistence is not often made explicit; many organisations cite a lack of tools that meet their specific need in their specific context; skill and competencies are mostly acquired in-house with some help from external consultants.
- **Key partners**: most providers apply LOGD in the context of existing peer networks; there is, as yet, little use of LOGD outside of those networks or by businesses.
- **Key activities**: in general, providers consider development and maintenance of LOGD services as part of their normal system maintenance and operational activities; few invest in promotional activities.
- **Cost structure**: given that many providers see LOGD activities as part of their core business, the study did not bring out the cost structure of the Linked Data activities alone as most providers do not yet separately account for this; where figures in terms of finances or staff resources were mentioned, these spanned a wide range depending on the approach taken.
- **Customer segments**: most cases showed either internal use or reuse in existing peer networks of government and non-government organisations; the study did not find much reuse of the LOGD by businesses.
- **Revenue streams**: the predominant revenue model is public funding, as part of the normal budgets of the organisations surveyed; in all cases, the data is provided free of charge; licences are either open or not explicitly defined.
- **Channels**: distribution channels include direct URI resolution and SPARQL endpoints. Bulk downloads are almost always offered; proprietary apps and Web applications are less common.
- **Customer relationships**: there is little branding or advertisement of LOGD services, and little user support; feedback is typically through informal communications as part of institutional collaborations.

We list below a number of enablers and roadblocks to the value proposition LOGD that were identified in the case studies. The enablers are:

- **Efficiency gains in data integration – the network effect:** providers are more likely to engage in LOGD activities if they can see an immediate benefit for themselves. In the same way that each new telephone added value to the existing ones at the birth of telecommunications, the addition of each new LOGD set adds value to those that are already published.
- **Forward-looking strategies**: as providers see the thrust of LOGD they may want to align themselves with modern techniques and technologies as a way to maintain their reputation as thought leaders in their domain.
- **Increased linking and integrated services**: providers who rely on connections with peer organisations, will value the possibilities for easier linking and increased interoperability that LOGD offers.
- **Ease of model updates**: LOGD makes future upgrades of data models much easier, for example to include new data or connect data from different sources together.
- **Ease of navigation**: URIs allow a 'follow-your-nose' navigation structure that provides better navigation through complex data.
- Open licensing and free access: LOGD considered in the study is mostly provided free of charge and under open licences which enables further use and reuse of data.
- **Enthusiasm from 'champions'**: the knowledge and enthusiasm of individuals in organisations who create awareness of possibilities and potential benefits help organisations to consider engaging in LOGD activities. When their efforts show real benefits, their employers are usually quick to offer support.
- **Emerging best practice guidance:** availability of guidelines and dissemination of best practices create common approaches and reduce risk in implementation by enabling organisations to learn from each other. Most of the organisations interviewed are convinced on the importance of sharing knowledge and experiences with others and contribute actively to the development of such best practices.

In addition, the following roadblocks were identified in the case studies:
- **Necessary investments**: as with all new technologies, LOGD requires investments in infrastructure, software and people. Not all organisations may be able to make such investments in a time of shrinking budgets and increased scrutiny.
- **Lack of necessary competencies**: not all organisations have the necessary skills, and specific training materials for a particular domain or application may not be readily available.
- **Perceived lack of tools**: some organisations currently develop their own tools, as there is a perceived lack of production-grade tooling. This is perhaps surprising given that Oracle, IBM and YarcData (part of Cray) are already among the companies offering high specification Linked Data systems. Additionally, the European Commission has funded in the context of the LOD2 project a number of open-source tools for Linked Data. Although the performance of RDF stores is a long way short of relational databases which are now highly optimised, resilient, production-grade systems, this should not necessarily be perceived as a barrier, as Linked Data services can run on top relational database environments.
- **Lack of service level guarantees**: the reuse of LOGD services by external third parties is hindered as providers do not yet give explicit service level guarantees. The case studies show that this is largely because the use of Linked Data is first and foremost for the

publisher's own benefit and the availability of the data for third parties is a side effect. Service Level Agreements do exist however in cases where the provision of the infrastructure is outsourced.

- **Missing, restrictive, or incompatible licences**: interviewees reported that missing, restrictive, or incompatible data licences continue to be a barrier to providing and consuming LOGD. It is not trivial to keep track of licence information for LOGD, especially when the ownership is not well defined or if data originates from different sources.
- **Surfeit of standard vocabularies**: many Linked Data applications are developed within a specific community with specific agreements using specific standards; although there are common standards like Dublin Core and FOAF, not all implementations use those in the same way giving rise to fragmentation that hinders wide interoperability.
- **The inertia of the status quo**: even more than other types of organisations, public sector bodies tend to favour incremental change such that new systems are seen as ways to replicate the same tasks as old ones. It was observed that in many cases Linked Data is seen as a more substantial change and therefore meets resistance. Additionally, as LOGD allows connections to be made and relations to be seen that were not visible in non-linked approaches, organisations see the technology as carrying a higher risk than more traditional approaches; uncertainties may lead to delays in adopting new approaches.

The detailed report of the study can be found at:

https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd

References

Berners-Lee, T. (2006, July 27). Linked Data. Retrieved December 02, 2013, from World Wide Web Consortium (W3C): http://www.w3.org/DesignIssues/LinkedData.html

Bizer, Heath, & Berners-Lee. (2009). Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, 1-22.

European Commission. (2010). European Interoperability Framework for European Public Services. Retrieved from http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

ISA. (2012). How Linked Data is transforming eGovernment. European Commission.

Official Journal of the European Union. (2003, November 17). Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:En:HTML

Official Journal of the European Union. (2013, June 27). Retrieved December 02, 2013, from EUROPA - European Union website, the official EU website: http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2013:175:SOM:EN:HTML

(2012). Open Data White Paper - Unleashing the Potential. Norwich: The Stationery Office. Retrieved from https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential

Open Knowledge Foundation. (2013). Open Definition. Retrieved December 02, 2013, from Open Definition: http://opendefinition.org/

PwC. (2014). Good practices for identifying high value datasets and engaging with re-users: the case of public tendering data. Retrieved from

http://www.w3.org/2013/share-psi/workshop/Timisoara/agenda

PwC. (2014). Do current licensing practices hinder the commercial reuse of open data? Retrieved from https://www.w3.org/2013/share-psi/wiki/images/5/5f/Bargiotti_Wyns.pdf.

PwC. (2014). Value-based prioritisation of Open Government Data investments. Retrieved from http://www.w3.org/2013/share-psi/workshop/samos/report.

The Open Knowledge Foundation (OKFN). (n.d.). Open Government Data. Retrieved April 22, 2014, from Welcome to Open Government Data: http://opengovernmentdata.org/

W3C. (2013). Linked data. Retrieved December 02, 2013, from World Wide Web Consortium (W3C): http://www.w3.org/standards/semanticweb/data